# An intelligent product suggestion algorithm using predictive analysis for personalized UI building

R. H. R. Perera

2019

# An intelligent product suggestion algorithm using predictive analysis for personalized UI

**A dissertation submitted for the Degree of Master of Computer Science**

**R. H. R. Perera**
**University of Colombo School of Computing**
**2019**

# Declaration

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name: R. H. R. Perera

Registration Number: 2016/MCS/073

Index Number: 16440734

Signature: Date:

This is to certify that this thesis is based on the work of

Mrs. R. H. R. Perera

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:

Supervisor Name: Dr. Lakshman Jayaratne

Signature: Date:

# Abstract

This research was done to introduce a personalized interface for item checkouts. Often we could observe lengthy queues in food outlets, supermarkets and other types of stores where the demand is high. Many studies are currently underway to find solutions to reduce the queue lengths and provide better service and satisfaction to the customers. This thesis reports a study of such scenario and how data analysis could provide a simple solution to the problem. As people expect more personalized experience nowadays, the solution for the problem was suggested as a personalized UI for the customers to select the items which they have purchased and create a checkout list all by himself/herself.

To generate personalized content, an algorithm was implemented to output the next purchasing item set using the historical purchasing records of the users. The algorithm uses a rule based approach with weighted ratings. Although collaborative method is a popular method in finding such results, in the studied scenario, it is not applicable as the store does not maintain a comprehensive user profiles or facilitate the uses to rate products. The research introduce a model named RFR-U model. The model uses the parameters; relevance, recency and frequency to determine the next purchasing item set. Since the algorithm could generate the results with-in seconds it could be used in real time applications. During the research a dashboard and a mobile application was also implemented to present and evaluate the results. According to the evaluation done, the algorithm accuracy level stands around 80% which is a fairly satisfactory rate compared to its simplicity

# Acknowledgements

First I would like to express my deep gratitude to my Supervisor Dr. Lakshman Jayaratne for the valuable advices and guidance given to me at all times, to make this research a success. I would also like to express my gratitude to the subject coordinator and the committee for the guidance and comments given.

I would also like to extend my thanks to Maheshika Wijesuriya, Kawmini Mahawattege, Esther Rijema, Shaheem Iqbal, Saminda de Silva and Chinthaka Sugathadasa for the continuous support extended to me in data gatherings and obtaining feedback.

Finally I would like to acknowledge the support provided by my family during the period of my final year project.

# Table of Contents

Declaration ........ i
Abstract ........ ii
Acknowledgements ........ iii
List of Figures ........ vi
List of Tables ........ vii
List of Abbreviations ........ viii
Chapter 01: Introduction ........ 1
1.1 Problem Definition ........ 1
1.2 Motivation ........ 2
1.3 Goals and Objectives ........ 4
1.4 Significance, Novelty and Criticality ........ 4
1.5 Scope ........ 5
1.6 Organization of the Thesis ........ 5
Chapter 02: Literature Review ........ 7
2.1 Introduction ........ 7
2.2 Predictive Analysis ........ 7
2.3 Product Recommending Systems ........ 8
2.3.1 Collaborative filtering ........ 8
2.3.2 Content-based filtering ........ 9
2.3.3 Other approaches ........ 9
Chapter 03: Methodology ........ 13
3.1 Introduction ........ 13
3.2 Problem Analysis ........ 13
3.3 Approach ........ 15
3.4 Design ........ 17
3.4.1 Design of the Algorithm ........ 17
3.4.2 Design of the System ........ 18
3.5 Methodology ........ 24

3.5.1 Data collection ........ 24
3.5.2 Data Analysis ........ 25
Chapter 04: Proposed Solution ........ 30
4.1 Introduction ........ 30
4.2 RFR-U Model ........ 30
4.2.1 Relevance ........ 31
4.2.2 Recency ........ 32
4.2.3 Frequency ........ 32
4.3 Variation of RFR-U model ........ 33
4.4 Dashboard ........ 34
4.5 Mobile Application ........ 34
Chapter 05: Evaluation and Results ........ 36
5.1 Introduction ........ 36
5.2 Quantitative Evaluation ........ 36
5.3 Results of the Quantitative Analysis ........ 40
5.4 Qualitative Evaluation ........ 42
5.5 Results of the Qualitative Analysis ........ 43
Chapter 06: Conclusion and Future work ........ 48
6.1 Introduction ........ 48
6.2 Critical evaluation of the project ........ 49
6.3 Future work ........ 50
References ........ 51
Appendices ........ 53

# List of Figures


Figure 1:High-level design of the algorithm ........ 18
Figure 2: Overall system diagram ........ 19
Figure 3: Architecture diagram of the tab based mobile application ........ 20
Figure 4: Module view of the mobile application ........ 20
Figure 5: Login Screen ........ 21
Figure 6: Home Screen ........ 21
Figure 7: After checkout. Checkout successful screen ........ 22
Figure 8: ER diagram of the backend database ........ 23
Figure 9: Purchases of user 1 ........ 25
Figure 10: Purchases of user 3 ........ 26
Figure 11: Time comparison of user 1 ........ 26
Figure 12: Time comparison of user 3 ........ 27
Figure 13: Date comparison of user 1 ........ 27
Figure 14: Date comparison of user 3 ........ 28
Figure 15: Day comparison of user 1 ........ 28
Figure 16: Day comparison user 3 ........ 29
Figure 17: Screenshot of the dashboard ........ 34
Figure 18: Login screen of the POC app ........ 35
Figure 19: Home screen of the POC app ........ 35
Figure 20: Answers for question 1 ........ 43
Figure 21: Answers for question 2 ........ 43
Figure 22: Answers for question 3 ........ 44
Figure 23: Answers to question 4 ........ 44
Figure 24: Answers for question 5 ........ 44
Figure 25: Answers for question 6 ........ 45
Figure 26: Answers for question 4 ........ 45
Figure 27: Answers to question 8 ........ 45
Figure 28: Answers to question 9 ........ 46
Figure 29: Answers for question 10 ........ 46
Figure 30: Answers for question 11 ........ 46
Figure 31: Answers for question 12 ........ 47

# List of Tables

*Table 1: Weight table for time factor* ................................................................ 31
*Table 2: Weight table of the recency parameter* ................................................................ 32
Table 3: Test results ................................................................ 40

# List of Abbreviations

RFR: Relevance, Frequency, Recency

RFR-U: Relevance, Frequency, Recency,User

RLF: Relevance, Frequency

RF: Frequency, Recency

RFM: Recency,Frequncy,Monotory

AI: Artificial Intelligence

UI: User Interface

POC: Proof of Concept

# Chapter 01

# Introduction

## 1.1 Problem Definition

Time has become a valuable entity in the modern world. Peoples' lives have become so busy that they dislike spending unnecessary time on their daily tasks. One such task is waiting in queues. Hence, many organizations have tried different solutions to reduce the waiting time in queues. As technology becomes an essential element in daily activities of humans, many organizations focus on finding technological solutions to this problem.

This thesis reports a study of such a problem and how data analytics can be used to minimize the effects of the problem. For this research, as the case study, the bakery outlets at a renowned supermarket chain in Sri Lanka has been considered.

The current system was first studied to see the causes and effects of the long waiting time in said outlets. One major reason causing the delay is the time taken to bill the products. Many a times the cashier is responsible for not only billing but packing as well. The delay is inversely proportionate to the experience level of the cashier. Billing task could be further broken down to punching items and payment processing.

The study also revealed that most of the customers who come to the bakery to purchase products are regular customers and most of them buy items on a daily basis. Generally, they visit the store to buy breakfast, lunch or an evening snack and often purchase the same set of items each day. Therefore, most likely there is a repetitive pattern in their daily purchasing behavior.

At present, a traditional POS system is used to check out the items. The store offers a loyalty program and most of the customers uses their loyalty number when purchasing items. Therefore, the data is recorded against a loyalty number, which can be used to identify buying patterns of each individual.

Considering the customers' and cashiers' behavior and tasks in the system, the problem could be defined as;

**The traditional method of item checkout has failed to serve the rising demands, leading to long queues in outlets and dissatisfied customers.**

Therefore, it is important to solve the problem of increasing queue lengths to increase the customer satisfaction and loyalty.

## 1.2 Motivation

As data becomes more and more important in the modern world, many businesses have adopted data analysis and big data concepts to simplify their daily processes and tasks [3]. Many organizations and researches have introduced smart applications of different scales to improve efficiency, customer satisfaction and return.

One such widely used area in data analysis is predictive analysis. Healthcare [2], retail, supply chain [1], weather forecasting [4] are few of the fields that benefits from predictive analysis.

This thesis discuss and evaluate how predictive analysis and recommending systems could be used to solve the identified problem. As the main cause of the problem is the delay in billing items, decreasing the time taken by the cashier to do any of the sub tasks which involves billing could reduce the total time taken by the cashier, leading to less service time and shorter queues. One of the sub tasks were identified as the punching of purchased items. If the task of punching items could be delegated to another person, the time to bill each customer could be drastically reduced.

The most suitable person to delegate the task was identified as the customer. If the customer could add the items he/she buys by him/herself then the cashier would only have the task of handling the cash. However, different varieties of customers visit the store and if an application is given to the customer to self-checkout, then the solution should be simple and attractive which can be used by any type of customers.

It has been identified by many researches that there are repetitive patterns in customer purchases and this could be used to simplify a self-checkout solution. [5] Even during the initial problem analysis it was observed that there is a repetitive pattern in each customer's purchases. Therefore, the purchasing patterns of the customers who visit to buy bakery items could be used to design and develop a simple and elegant self-checkout application and this thesis reports on how the repetitive patterns could be used design such an application.

A key factor in a self-checkout application should be simplification. A single bakery outlet sells a wide range of items and a typical application would list all the items. This would be a complex design since the customers will have to go through the whole list to add items. This requires a considerable effort and after sometime the customers would stop using the self-checkout function and the problem will rise again.

Therefore, it is important that the solution which is given to the customer is user friendly and simple so that they would continue to use it. Hence rather than showing the complete set of items, if the solution could predict next purchasing item list, the priority could be given to those items in displaying the list so that the customer would not have to search for the items he/she buys.

By reviewing the literature, it was found that many researches have used predictive analysis methods to come up with such predictions. One research has been done to recommend products to customers using the resent purchases of the customers [6]. It has used both customer information and product information to calculate results. There are many other similar researches carried out to identify buying patterns and predict next buying item lists. A detail study of such researches are mentioned in the literature review chapter.

## 1.3 Goals and Objectives

The main objective of the research is to design and implement an efficient algorithm to find a list of potential items that a customer would buy at a given time. The other objectives of the research includes;

- Identifying buying patterns of customers,
- Suggesting promotions or new items according to the patterns identified,
- Evaluating related work done by other researchers to find out how they have addressed similar issues.
- Evaluating the research qualitatively and quantitatively.

## 1.4 Significance, Novelty and Criticality

The most significant attribute of the algorithm is the real time data analysis. Time factor is extremely critical as the algorithm will be used for self-checkout applications and the results should be populated within seconds. Therefore, it is important to choose a correct portion of data for the analysis. Overloading the data may cause delays in executing the algorithm and insufficient data will predict incorrect results. Many algorithms exist to extract patterns and make predictions offline. But, the reported research focusses on producing results real time.

In addition, many predictive analysis algorithms use collaborative based techniques and the reported research is done using a content and rule based method.

## 1.5 Scope

Research similar to the area of interest was studied and the thesis describes how similar problems have been solved. The literature was critically evaluated to study the techniques and methods used to develop algorithms to generate results. Further, the evaluation mechanisms were studied to check if any improvements can be done to the planned evaluation method. This thesis elaborates a detail study of the existing literature on similar topics, suitable techniques and methods identified to be used for the proposed research.

The dataset were analyzed using graphs and simple statistics to identify variables that derives the buying patterns. The findings and the background study is also present in the thesis. Using the techniques, variables and other important factors identified during the background study and the initial analysis of the data set, an algorithm was written to get the possible item set that a given customer would buy at the given time and date (personalized suggestions).

A tab based Android mobile application was developed to evaluate the designed algorithm. The accuracy of the algorithm was tested using random bill checking and a detail elaboration on the evaluation is mentioned in the evaluation chapter. The application was given to few users in the system and the overall experience of the users and the satisfaction was tested using the feedback given from the users to check if the research problem mentioned in the thesis were solved using the hypothesis.

## 1.6 Organization of the Thesis

This thesis reports the design and implementation of an algorithm which predicts the next item set of a customer. The organization of the thesis is as follows.

The chapter two reports the literature review. Predictive analysis is a branch of data analytics and recommending systems are one of the uses of predictive analysis. Product recommending techniques can be mainly classified into three categories; collaborative filtering, content-based filtering and hybrid approaches. This chapter analyzes the methods available to find the most suitable method for the problem analyzed.

Chapter three is the methodology chapter. This chapter contains a detail problem analysis including the causes of long queues and customer behavior. The hypothesis for the research is also present in the chapter and what the solution could address. The chapter also includes the design of the algorithm and the system. During the analysis it was identified that the three factors, time, date and frequency contributes significantly towards the patterns of purchase and the chapter elaborates the importance of those factors.

Chapter four elaborates the proposed solution. The three factors identified has been used to create the RFR-U model which can predict the next purchasing item set. The model uses Recency, frequency and relevance of the records to calculate the probability of purchasing item set. The chapter explain few variations of the model as well.

Chapter five contains a detail evaluation of the research. The research has been evaluated in both quantitative and qualitative approaches. This chapter also includes the results of the evaluation. The accuracy of the algorithm stands above 80% and the feedback obtained from the users are mostly positive.

Chapter six states the conclusion. All the finding and results are summarized in this chapter. The accuracy of the algorithm stands above 80%. The chapter includes the critical evaluation of the research and the future enhancements that could be done.

***

# Chapter 02

# Literature Review

## 2.1 Introduction

As digitalization happens in almost every organization and businesses, thousands of data is gathered through each system every second. Many organizations are now trying to use these data to implement and innovate new systems. Also many research in carried out to use these data to make systems more efficient and interactive. The purpose of this chapter is to introduce and evaluate the research done in the area of predictive analysis and how it can be used in product recommending systems.

Product recommending methods such as content-based filtering, collaborative filtering, hybrid methods, rule based methods have been studied to determine the best approach to be used for the reported research.

## 2.2 Predictive Analysis

*"Predictive analytics encompasses a variety of statistical techniques from data mining, predictive modelling, and machine learning, that analyze current and historical facts to make predictions about future or otherwise unknown events."[7] – Wikipedia*

Predictive analysis is a branch of the advanced analytics which is used to make predictions about the future. Many areas has been benefited from predictive analysis, such as weather forecasting, healthcare, supply chain and automobile.

The research reported in the thesis focuses on the retail sector. Predictive analysis has been widely used in the retail sector to solve many problems. One such scenario is identifying consumer buying patterns.

Many researches are done to predict and identify consumer buying patterns. The results of these are widely used in decision making, planning, sales and marketing etc. But the main focus of the reported research is to use identified buying patterns to recommend or suggest products for a self-checkout application.

Product recommendation algorithms can be mainly classified into three categories. Content-based filtering, Collaborative filtering and the hybrid approach [14]. Some researches suggests another category called Rule based filtering [15].

Content-based filtering methods are based on the users' preferences in the past. These algorithms analyses the history of the users' behavior, to come up with suggestions. Collaborative filtering techniques analyzes large amount of data and find similarity between users. The products are suggested according to the similarity of user preferences. Therefore the recommendation is not only based on the individual user itself. Hybrid approach is a combination of the content-based method and collaborative method.

Another researcher mentions another category named rule based. *"Rule-based approach is a simple but popular way of recommendations. Rules are usually derived from database of previous transactions."[15]*

## 2.3 Product Recommending Systems

### 2.3.1 Collaborative filtering

The most commonly used product recommending approach is the collaborative model. "*It aims to identify customers whose interests are similar to those of the current customer, and recommend products that similar customers have liked" [17]*. Some approaches in this category are k-nearestneighbor, matrix factorization and semi-supervised learning. [16]. Many techniques such as user profiles, ratings given by users are used to measure similarities between users and suggest products.

Although collaborative filtering is the most commonly used approach this method is not suitable for the reported research as the store does not maintain a comprehensive user profile and the store does not have a rating mechanism on the products. With the data available it will be complex to find similarities there by making the algorithm complex and time consuming to execute.

### 2.3.2 Content-based filtering

*"Content-based filtering methods are based on a description of the item and a profile of the user's preferences. These algorithms try to recommend items that are similar to those that a user liked in the past (or is examining in the present)" Wikipedia*

The content-based recommendation systems are mainly used to recommend documents, web pages, publications, jokes or news. This method could be used to suggest a new item based on the past interests.

In the research done by Haiyun Lu consecutive subsequences are used to derive patterns. This uses principle of sequential method to form consecutive subsequences. The consecutiveness ensures the recency of extracted patterns. Then it builds a conditional probability model to make recommendations. [15]

### 2.3.3 Other approaches

One research has used customer segmentation to predict and recommend products. The paper talks about the drawbacks of the collaborative filtering and recommends segmentation intend. It identifies RFM model and an enhanced version of RFM model as an effective model for user segmentation.

RFM model has been proposed by Hughes in 1994 [8]. It proposes three behavioral variables, namely Recency (R) Frequency (F) and Monetary (M). Recency denotes the latest purchases, frequency denotes the total number of purchases and monetary denotes the monetary value spent during one period [9]. However the paper mentions about the drawbacks of the model as well and mentions about a revised model proposed by Macus. He used the number of purchases (F) and the average purchase amount (A) construct two-dimensional matrix model based on CLV (customer lifetime value) to correct the RFM method [10].

Customer segmentation methods can be classified into two categories, i.e. market research or data mining methodologies and It has used neural networks and fuzzy logic to come up with a solution. The case study for the reported release being bakery items purchases, it is highly likely that the results depends on the recent purchases than purchases made months or years back. Therefore the two behavioral variable**, recency** and **frequency** in the RFM model can be considered as two important variables in the reported research as well.

Another research paper studied was the paper titled, “An Intelligent Product Recommendation Model to Reflect the Recent Purchasing Patterns of Customers” [6]. The research uses the recent purchasing patterns to recommend product to customers. It has considered both customer information and product information for the solution. The research has used various data mining classifiers such as the decision tree, neural network, support vector machine, random forest, rotation forest, sliding-window scheme for the recommendation model recommendation model.

Another research uses a combination of attribute based approach and a sequential based approach. It makes a personal preference matrix for each user to generate results. In the sequential based approach it uses weighted association rules to identify latest patterns. It suggests that the purchasing processes usually have some time-dependency relationship and is repeatable and also periodical. [16] This is similar to what was observed during initial problem analysis of the reported research

Rule based approach is also a simple approach of recommendation. Rules are usually derived from database of previous transactions. This uses data mining techniques such as web usage mining, decision tree induction to find association rules. [15]

## 2.4 Real time Predictions

Although there exist many data analysis and prediction algorithm, most of them does not provide real time results. One key attribute in the reported research is producing results real-time. Therefore it is important to review and evaluate research done on producing results real time.

## 2.5 General Data Protection Regulation (GDPR)

As the research is done using customer data, the regulations related to customer information should also be studied. GDPR is one such regulation.

General Data Protection Regulation (GDPR) is a legislation approved by European Union Parliament on 14 April 2016. However, the law was enforced on 25th March 2018. Main objective of the GDPR is to protect personal identification information (PII).

Under this organizations must ensure that personal data is gathered legally and also under strict conditions. Organizations who collect and manage the data will be obliged to protect it from misuse and exploitation, as well as to respect the rights of data owners. If any breach on the above conditions the Organization will be penalized under the GDPR Act. [13]

This new EU framework applies to organizations in all member-states and has implications for businesses and individuals across Europe, and beyond. The rules don’t apply to data processed by an individual for purely personal reasons or for activities carried out in one's home, provided there is no connection to a professional or commercial activity. Also GDPR applies to the processing of personal data wholly or partly by automated means as well as to non-automated processing, if it is part of a structured filing system. [13]

GDPR can be a huge limitation for the Big Data Analytics related projects. Accuracy of the Big Data Analytics modules will be affected by this regulation. If the Analytics involved in profiling the customers then the business should take the individual consent of the customers before doing so. Number of data to work with for data scientists will be reduced with the GDPR.

There are no Silver bullets for the GDPR. Only way of achieving the Big Data Analytics with GDPR is being Compliant with the regulation.

***

# Chapter 03

# Methodology

## 3.1 Introduction

This chapter includes a detail analysis of the problem and its causes. This also elaborates on how predictive analysis can used to solve the problem. The chapter also discuss on how the data was gathered and analyzed. Graphs and simple queries were used to analyze the data set to find the deciding factors of item of purchase. During the analysis it was identified that time, date and frequency contributes majorly towards the pattern of the purchase. According to the analysis the algorithm was designed to find the most probable item set that a user might purchase on a given time or date. The chapter includes the systems design diagrams and architecture as well.

## 3.2 Problem Analysis

Below mentioned is the identified problem during the initial stages of the research. According to the identified problem the problem analysis were carried out during the research.

**"The traditional method of item checkout has failed to serve the rising demands, leading to long queues in outlets and dissatisfied customers."**

Long queues are formed mostly during peak hours because a large number of customers enter into the system and the service rate is not the same as the arriving rate. Since the arrivals could not be controlled, what could be controlled and improved is the service.

After close observation of the system the below causes was found for the delay at the cashier.

- Customers coming to bill the items sometimes spend a lot of time in the cashier trying to decide what to purchase
- The cashier fails to understand the customer needs at the first time that he or she has to repeat the order again
- When foreign nationals comes, sometimes they fail to understand their requirements due to language and pronunciation barriers
- Cashiers get distracted
- Payments also delays billing; Credit/Debit card not working etc.
- If the cashier is a trainee he/she spends time on finding the product codes
- Some people buy a long list of items that it takes a while to bill the purchase and the customers who are next in the line have to wait till it’s over. Some stores offers a express counter for few item purchases but in this case no such counters are available

It is important to study customers’ behavior to derive a better solution. Therefore during the analysis the customer and his/her behavior was closely observed. The finding of the observation is as follows;

- Customers are of different types in terms of behavior. Education, job roles etc.
- Most of the customers are regular customers
- Customers enter into system for different purposes.
    - To buy a main meal
    - To buy a snack
    - To have some desserts
    - To drink something refreshing
    - To experiment new additions
- Customer’s purchases differ according to the date, day of the week, day of the month, time of the day, weather etc.
- Most of the regular customers purchase the same set of items with minor differences.

It is highly important for a business to know their customers. This would assist the organization or the business to maintain a good relationship with the customer and also to stop the customer from going to competitors. It would also help to individually treat the customer according to their likes and dislikes.

Nowadays almost every retail store have loyalty schemes and all the purchases of the customers can be found in the digital formats. Even In the case study the store offers a loyalty scheme and the purchases are recorded against the customer information. This means that by analyzing the purchasing data of the customer, the organization could get to know the customer better.

## 3.3 Approach

Since the traditional methods have failed to cater the rising demands and expectations of the customer, alternative methods should be studied to resolve the problem. As digitalization is on the rise in the global market and all the organization are turning toward digitalization, it is reasonable to look at the trending digitalization concepts to derive a solution. Data analysis, artificial intelligence based solutions, machine learning, and mobility are some of those popular topics in the digitalization stack.

As mentioned in the literature review data analytics has evolved greatly in past years and has been used in decision making, predicting etc. In the literature review it was revealed that some organizations have used data analysis and predictive analysis techniques to implement self-checkout applications. After studying the similar systems and researches it was decided to follow a similar approach to solve the mentioned research problem.

Today, one of the hot topics in the industry is, smart devices, smart application, AI embedded UI etc. Therefore traditional application with static user interfaces will no longer impress the customers nor will it bring much benefit to the organization. Therefore it is important to use techniques to make the solution smarter. In the reported scenario, being smart could be defined as, knowing what the customer is about to purchase. If that could be achieved, the application could prioritize the list of item set when prompting the complete item set. This makes the self-checkout much easier and faster.

Therefore the hypothesis for the research can be stated as follows;

**“Factors such as Time, date affects the users purchase and these variables could be used to develop an algorithm to predict the next purchasing item list of the user and it could be used for intelligent UI building for a self-checkout application”**

How could the said smartness can be achieved was the next issue to be addressed and research on. As mentioned earlier in the thesis, the store maintains a loyalty scheme and the customer purchases are available in digital format. These data was used to identify different buying patterns of individual customers, which at the end could be used to predict the next purchasing item set.

After obtaining the necessary data, the data set was studied to identify key variables which affects the patterns of the customers. It was decided to implement an algorithm to generate the next purchasing item set. The results of the algorithm was used in a self-checkout application.

The important point to note here is that using the results of the algorithm the application could function as a smart application, because it does not display a static set of items. Instead the application is able to display the predicted set of items for the particular customer, making the check-out easier and efferent.

A self-checkout could address the following identified root causes;

- Delay in entering the items to the system: Cashier need not to enter each individual item to the system. By scanning the code, the list of items will be automatically populated.
- Time taken to decide on what to buy: The customers could take their own time as the time taken by him/herself will not affect the other person
- Language barriers and misunderstandings: Since the customer has the facility to add the item to the bill, the necessity of communication decreases.

## 3.4 Design

The design of the research can be broken into two major areas. Design of the algorithm and the design of the application.

### 3.4.1 Design of the Algorithm

According to the problem analysis the following variables were identified as potential deciding factors for a purchase.

1. Time of purchase
2. Day of purchase
3. Ongoing promotions
4. Customer preference of taste
5. Weather

Out of these the first three factors were analyzed to check if the relationship with the purchase is significant. The factors analyzed are Time, date and frequency. A detail analysis of the three factors are present in the data analysis section of this thesis.

Since the data should be generated real time, the most important attribute of the algorithm will be performance, meaning the algorithm should be designed in such a way that the results will be generated in few seconds. Therefore, to predict the results only the recent purchases have been considered. Therefore, it is important that the algorithm capture the patterns with a limited amount of data.

As the first approach a matrix method was used. However, the large number of different items made it more complex to determine the next purchasing item set. Therefore a different method had to be tested to generate the list.

The second approach used is a combination of rule based and content-based models. Different rules related to identified variables; time, date, day and frequency was experimented to determine a suitable filtering approach. As found during the literature survey, some of the researchers have used rating methods to prioritize items. Therefore different rating methods was also examined during the experiments. After evaluating different approaches an algorithm was developed and the high-level design of the algorithm is as follows

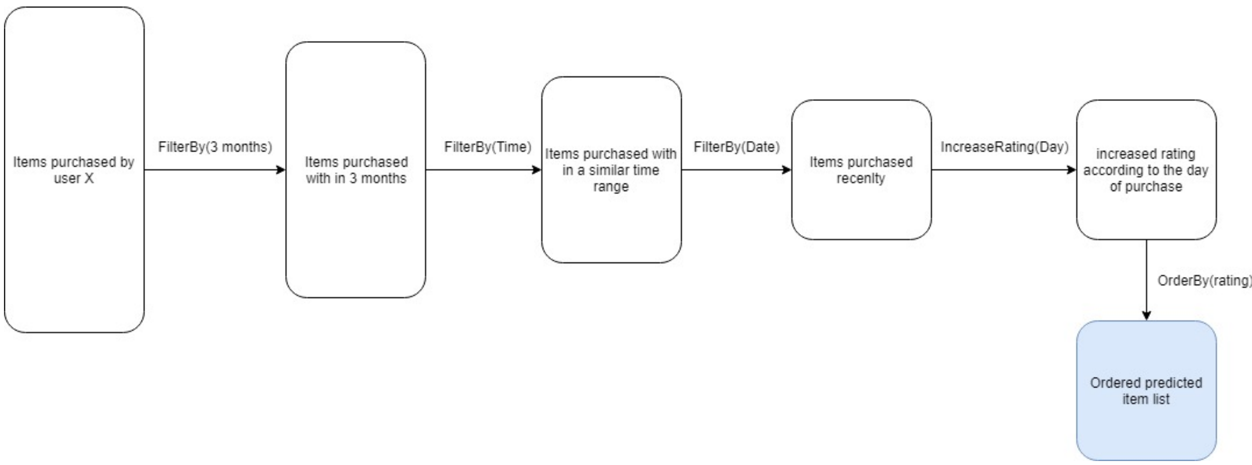



*Figure 1:High-level design of the algorithm*

The algorithm uses historical data of three months from the date of purchase to generate the results. A simple rule based method is used to produce results. Thus makes the algorithm simple and fast.

The algorithm will calculate a rating for each items according to the time, date, frequency and the day factors. The comparisons are made between the current purchase date and time (hereafter referred as "purchasing date") and the historical item purchased date and time. (Hereafter refereed as 'item purchased date') The calculated ratings will be ordered and percentage of the rating will be calculated for each item.

### 3.4.2 Design of the System

The system consists of two main parts.

- Tab based mobile application
- Dashboard to illustrate results
- Backend which host the data and the algorithm

### System Overview Diagram

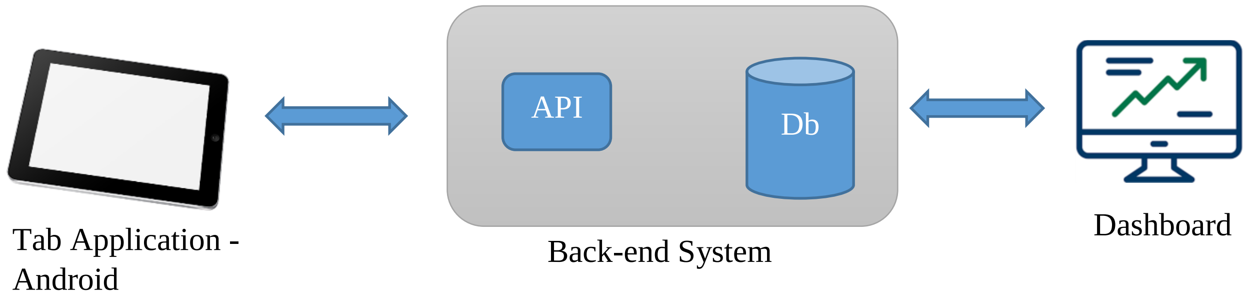



*Figure 2: Overall system diagram*

### Tab based mobile solution

The tab based application will have to following features;

1. Login with the user id
2. Display list of predicted items according to its priority.
3. Select items for the food basket (i.e. shopping cart) to be purchased
4. Select quantities
5. Checkout from food basket
6. Generate an unique code for billing
7. Suggest items of similar interest. (Which is not in the predicted list)

**Architecture diagram of the mobile solution**

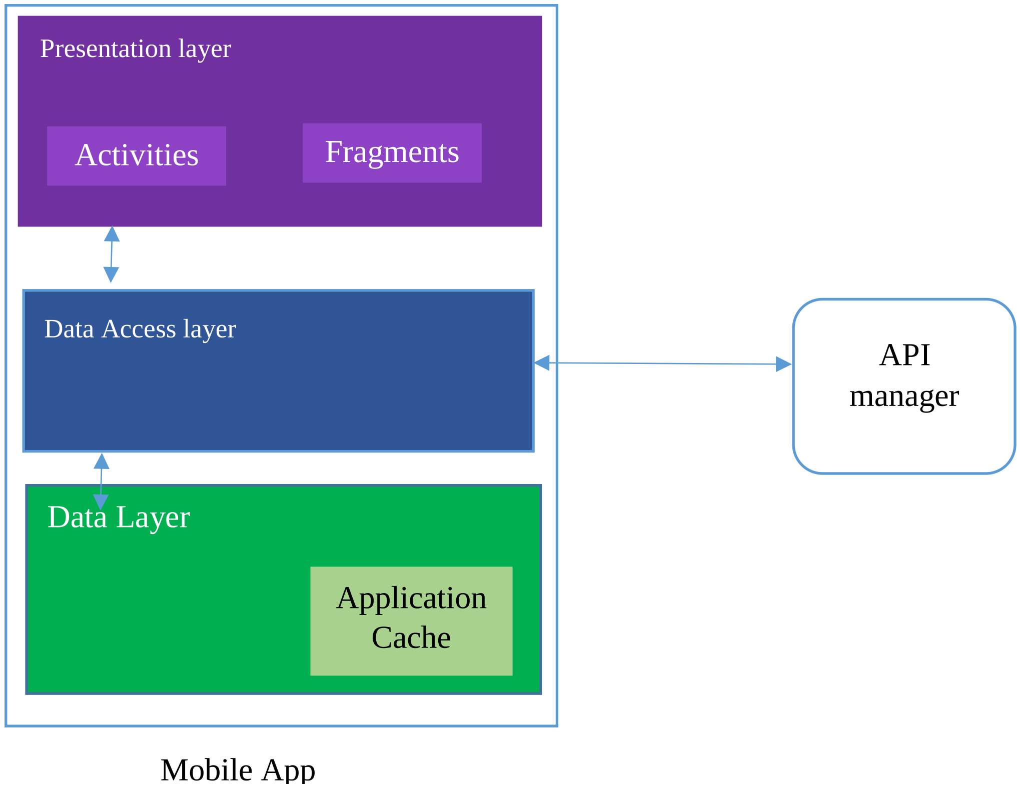



*Figure 3: Architecture diagram of the tab based mobile application*

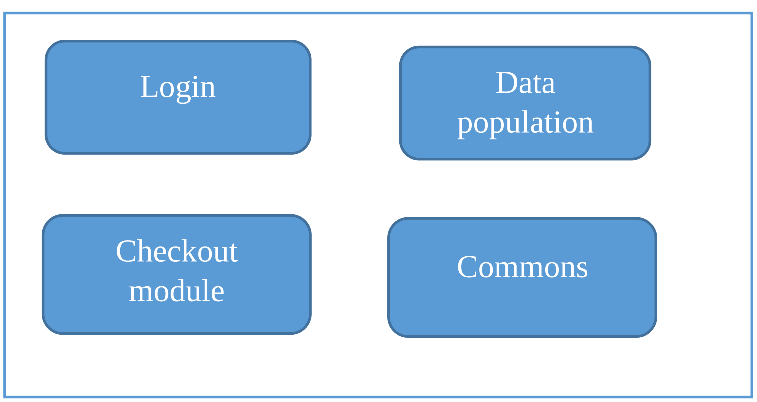



*Figure 4: Module view of the mobile application*

**Mockup screens of the mobile solution**

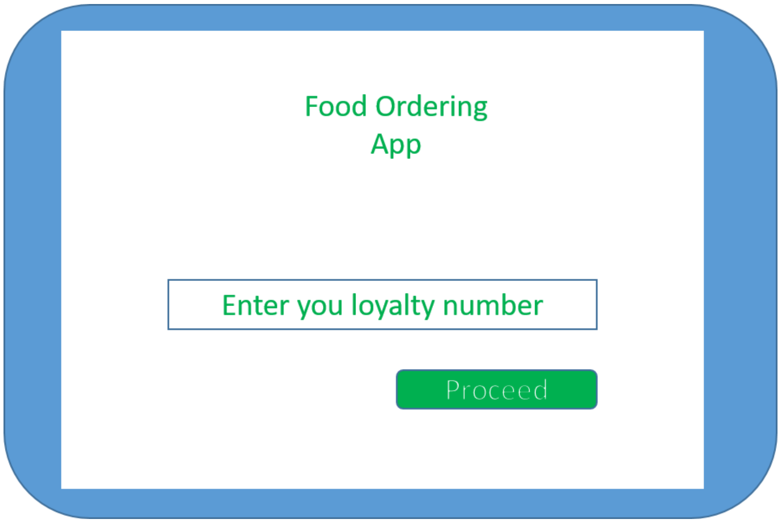



*Figure 5: Login Screen*

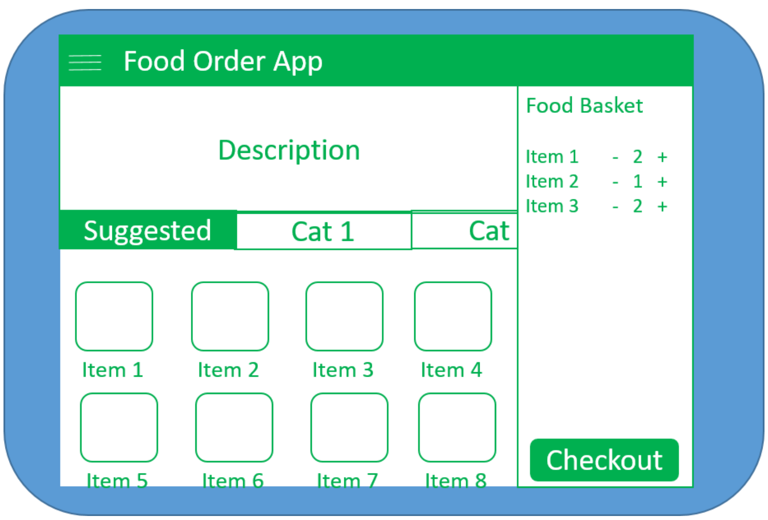



*Figure 6: Home Screen*

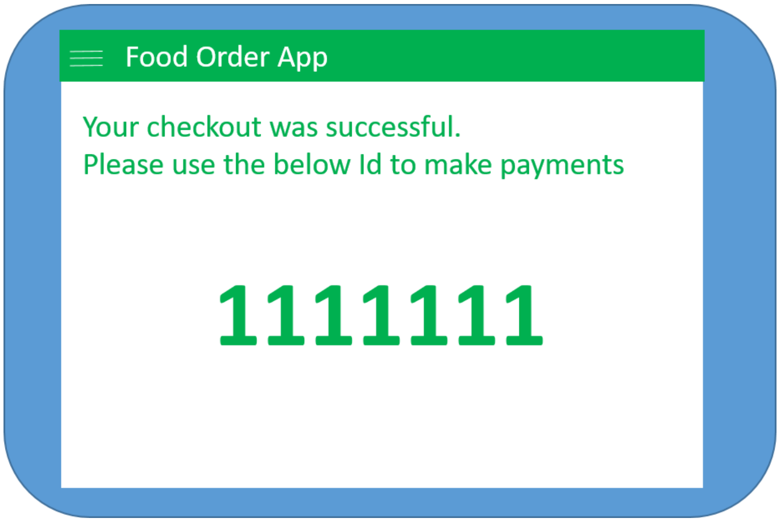



*Figure 7: After checkout. Checkout successful screen*

*Backend Solution*

The backend will host the algorithm, the database and the set of APIs. The tab application communicates with the backend using the API manager. Also a dashboard was created to present the results of the algorithm

The following APIs will be available in the backend

1. Login API
2. Get item set API
3. Generate code API

The following ER diagram illustrates the database design

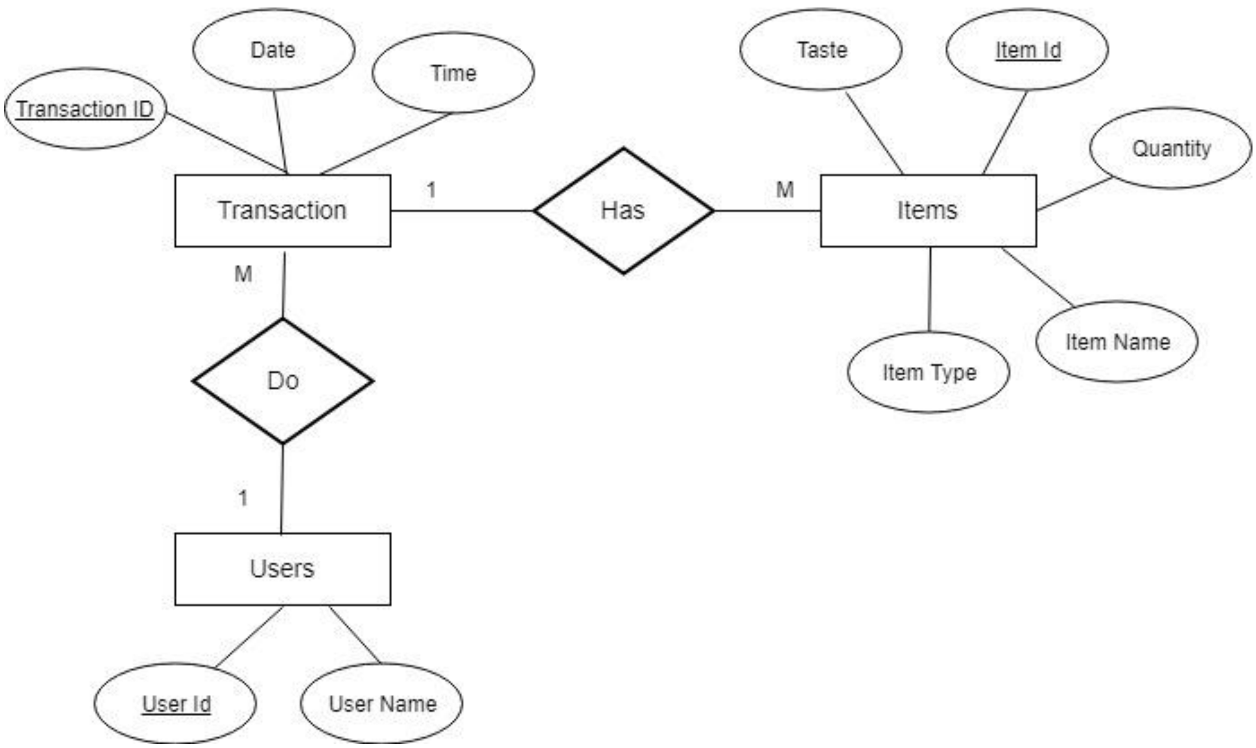



*Figure 8: ER diagram of the backend database*

## 3.5 Methodology

### 3.5.1 Data collection

The data was collected from a set of sample users over a period of six months and their bakery item purchases were closely monitored. Their bills were collected whenever possible. The data that was collected has been entered to a database. Only the relevant data was picked form the bills and entered into the database to ensure minimal noise.

As proposed the data set from the desired outlet could not be obtained due to a sign-off of a compliance named GDPR.

“The General Data Protection Regulation (GDPR) is a regulation in EU law on data protection and privacy for all individuals within the European Union (EU) and the European Economic Area (EEA). It also addresses the export of personal data outside the EU and EEA areas. The GDPR aims primarily to give control to individuals over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU.”[7]

Since the citizens of Europe countries purchase items from the outlet, if they fail to maintain the compliances the organization will have face heavy fines. Therefore the organization was not able to share the personal information of the customers with a third party. However necessary data was collected through colleagues and friends who regularly purchase data from the outlet. As the method used in the research is a content-based method and the item set is determined only using the historical data of the particular user, the data obtained through manual collection is still valid to develop the algorithm.

Many researches have been done on the impact of GDPR on data science research and development. Few incites of those researches are mentioned in the literature review section.

### 3.5.2 Data Analysis

Since the data was entered manually, no data cleaning or data translation was required during the analysis. It was assumed that the items purchased differ according to the data and time of purchase. To check the hypothesis a correlation analysis were carried out.

Two users; user 1 and user 3 was taken as sample users to evaluate the hypothesis. The frequency count of each item purchase between the months of September and December were considered during the study. The total number of unique items purchased for user 1 and user 3 is illustrated below. The number of items purchased by each user counts to a large number.

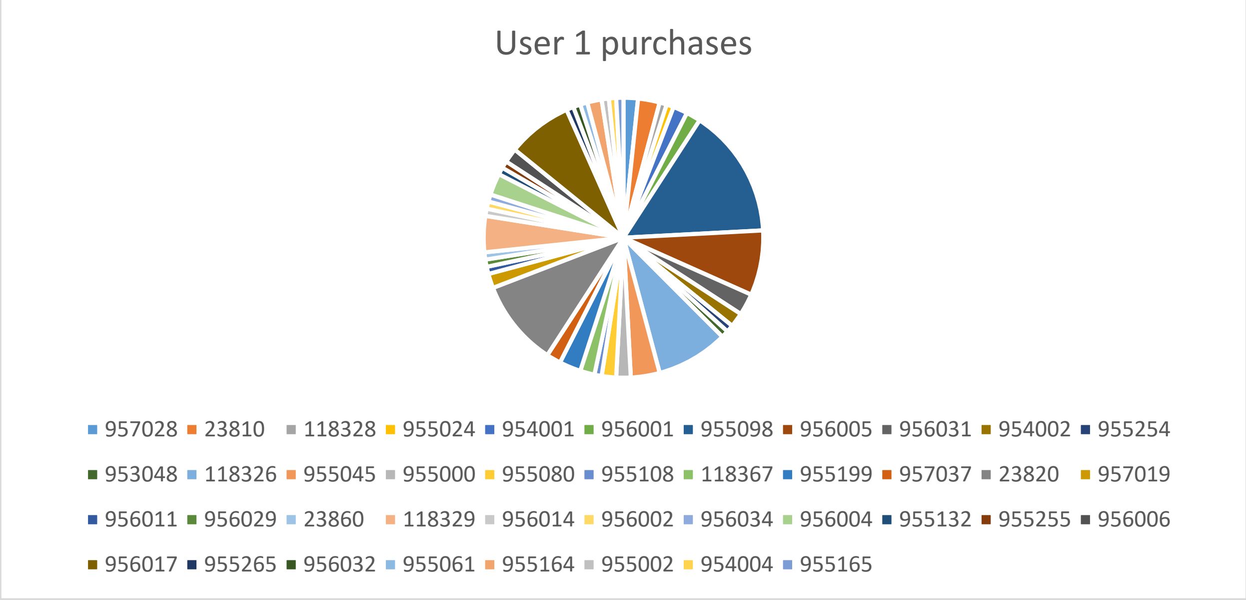



*Figure 9: Purchases of user 1*

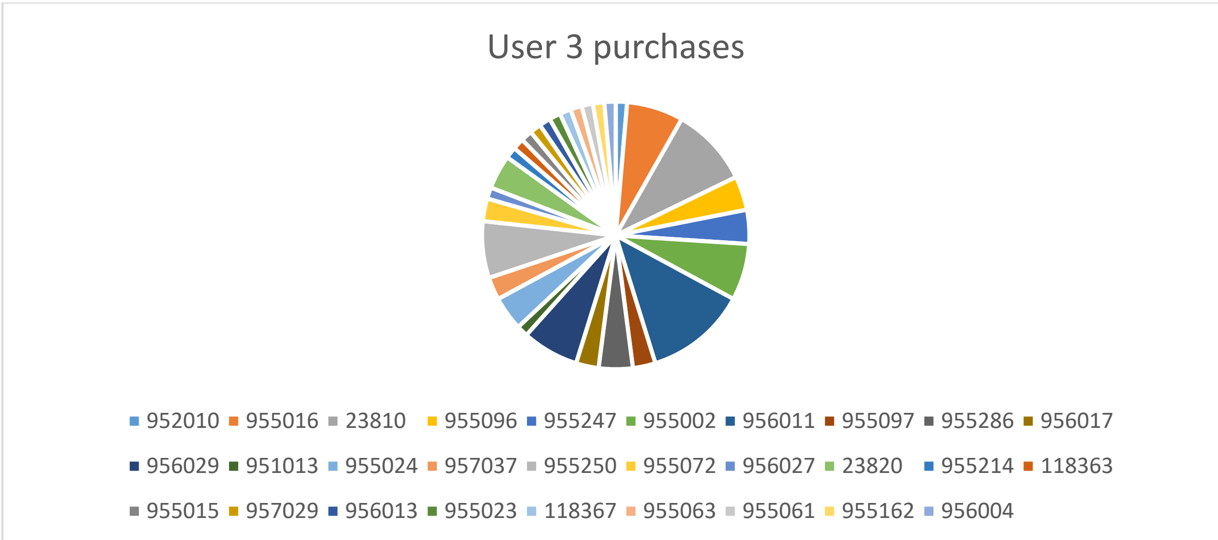
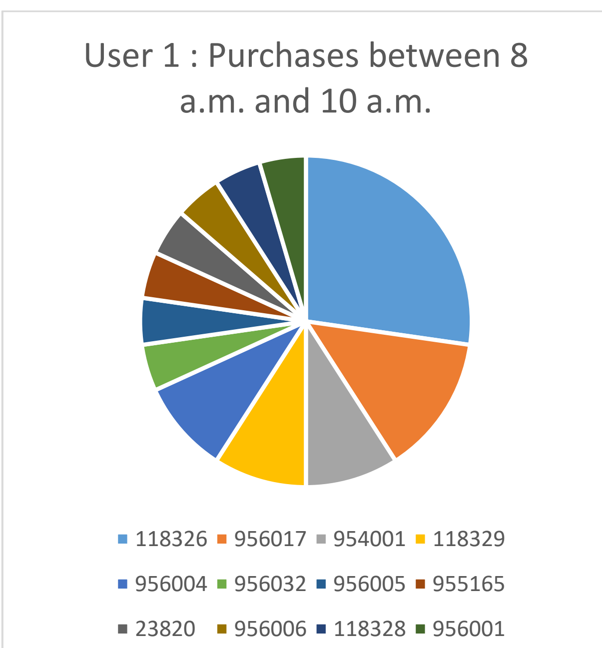



*Figure 10: Purchases of user 3*

**Time of Purchase**

The result set were filtered by different time periods. The following graphs illustrate the result sets. The filtering reduces the item count by approximately 50%. Therefore it is clear that the time of purchase impacts the purchasing result.



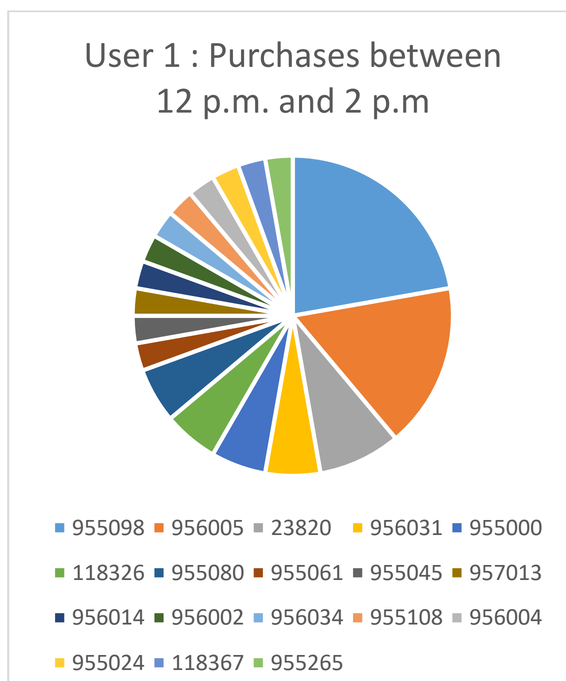
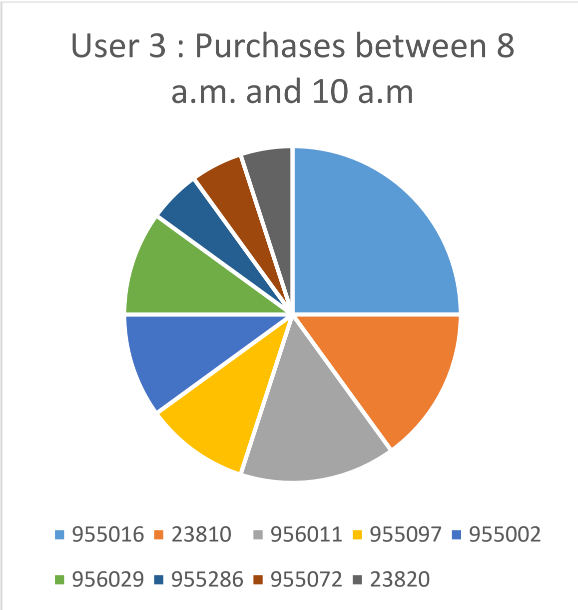




*Figure 11: Time comparison of user 1*



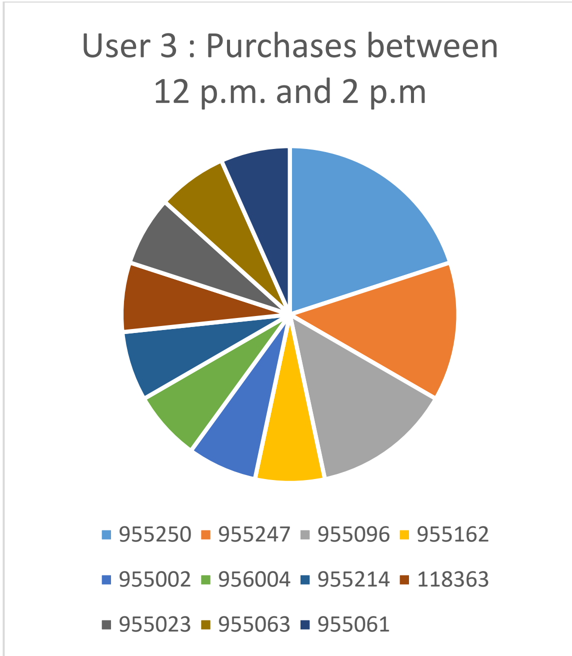
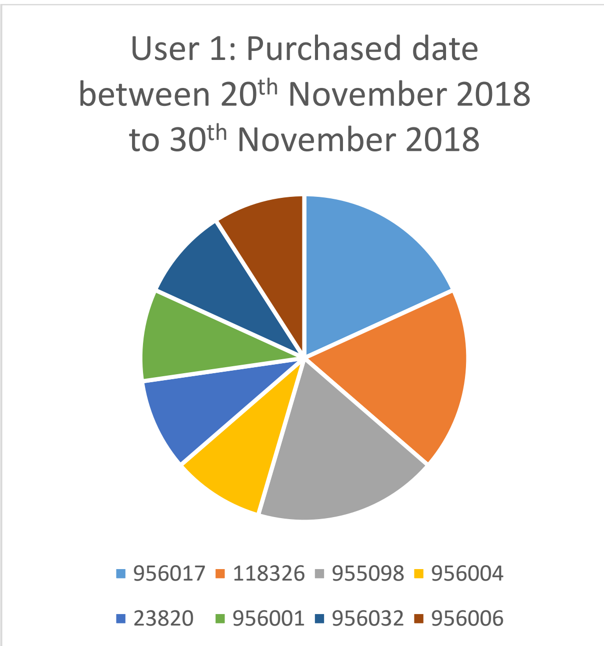




*Figure 12: Time comparison of user 3*

**Ongoing promotions/ Recency**

This could be monitored using the recency of the purchase. The following charts illustrates the results after filtering. It is clear that the recency also has a significant impact to the purchase.



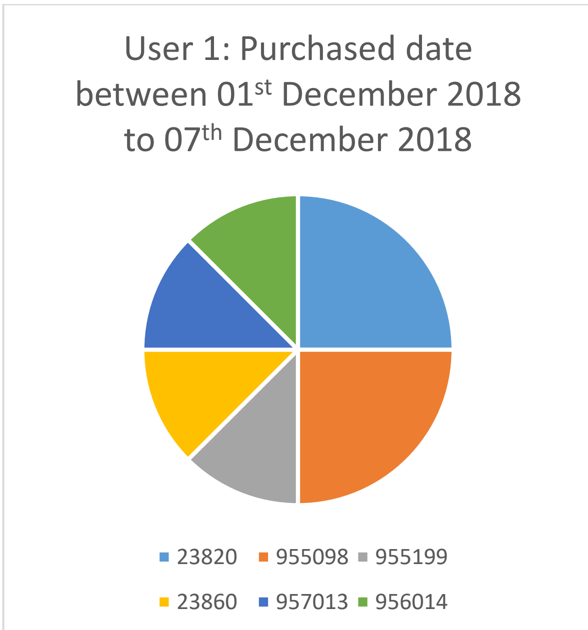




*Figure 13: Date comparison of user 1*

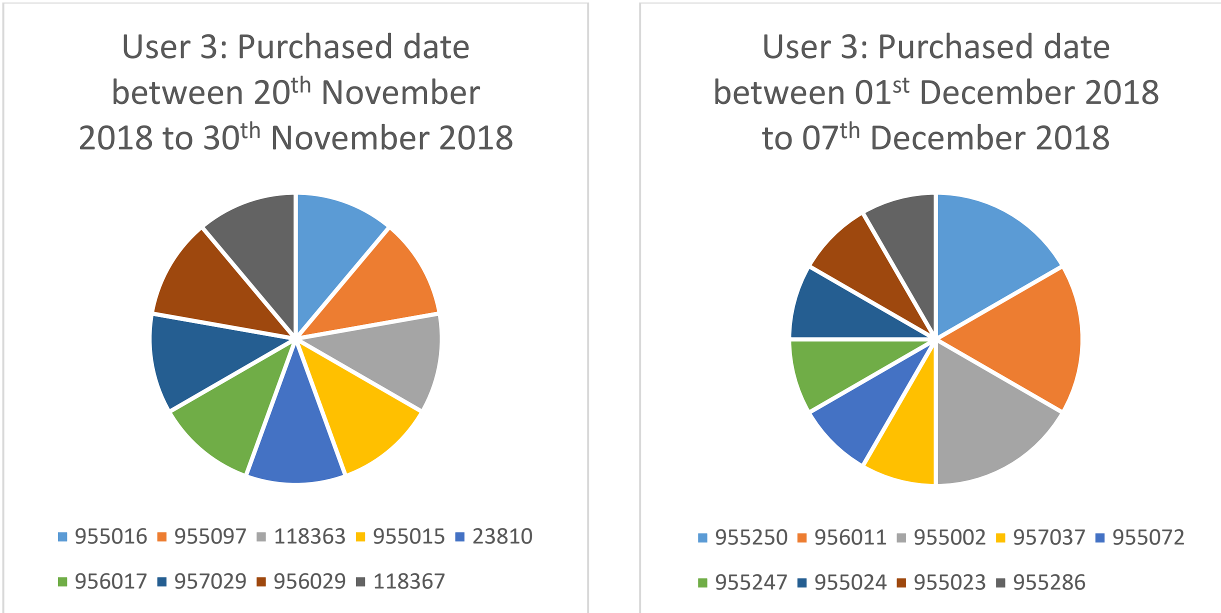



*Figure 14: Date comparison of user 3*

**Day of the week**

The results were also filtered using the day of the week to check if there is a significant impact. Even though the impact is less comparing to the other factors, the impact can be still considered as significant.

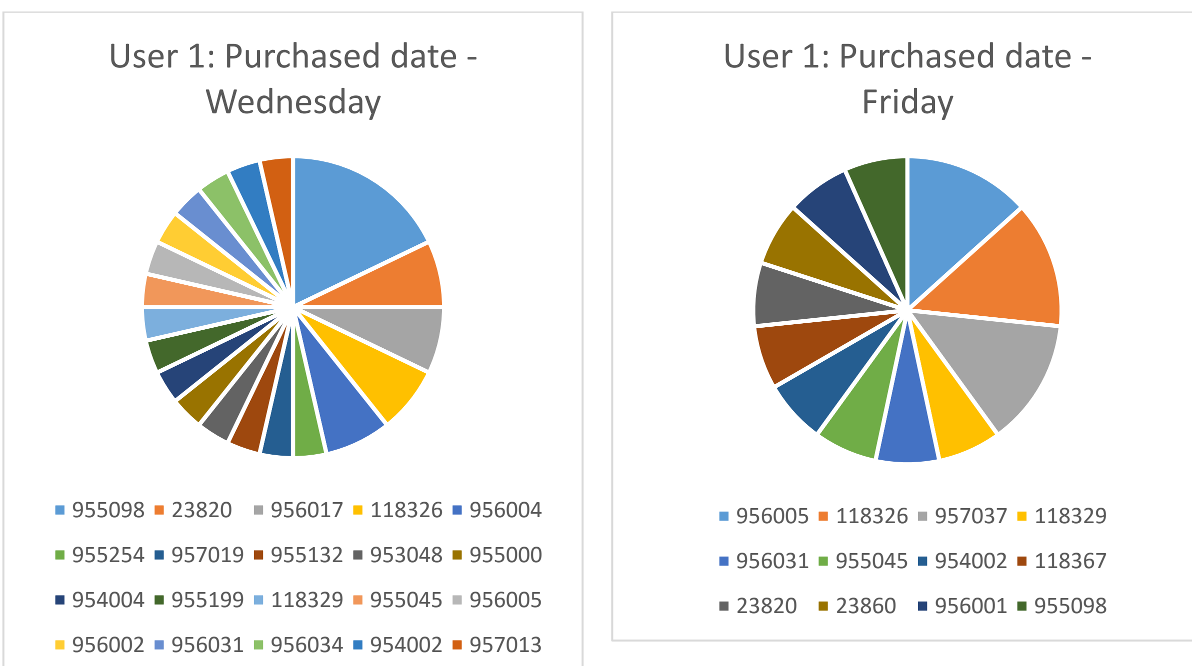



*Figure 15: Day comparison of user 1*

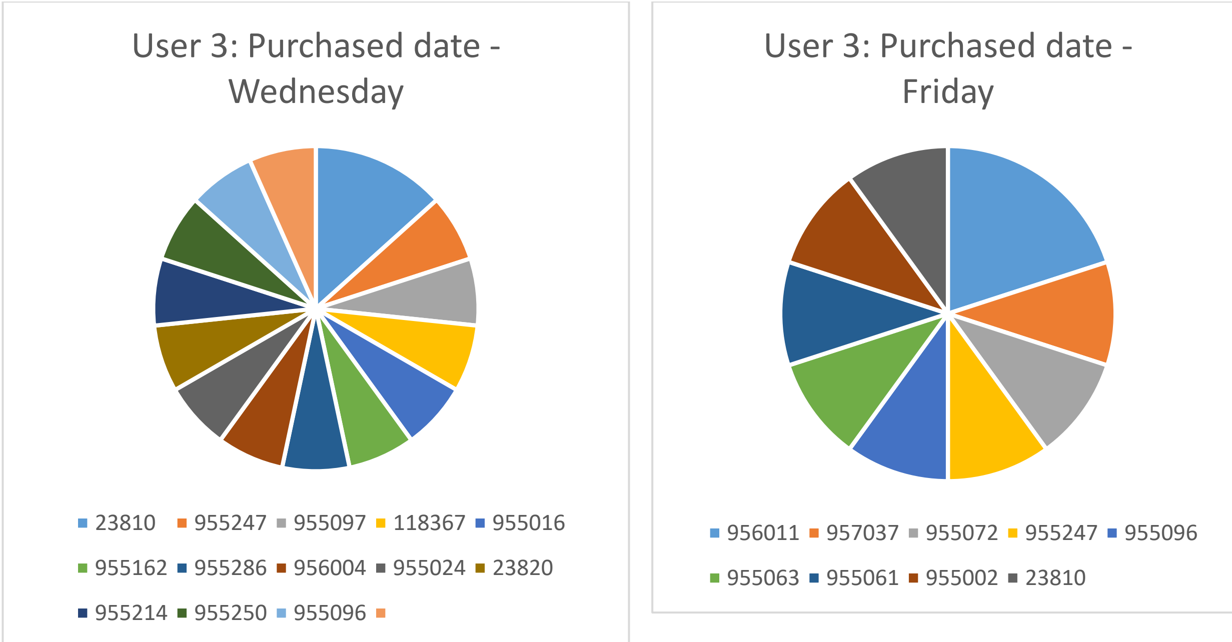



*Figure 16: Day comparison user 3*

After a detail analysis of the factors the following variables were identified as key variables of defining customer purchasing patterns,

1. Date of purchase
2. Time of purchase
3. Day of purchase
4. Frequency of the items

***

# Chapter 04

# Proposed Solution

## 4.1 Introduction

After careful evaluation on the literature and the analysis of the key variables, a basic model for the algorithm was developed. The model is based on the following factors, Recency, Frequency and Relevance. Also the model uses the history of the intended user. Hence the model is named as RFM-U model.

For the back-end, a dashboard was created to view the results of the RFR-U model and its variations. In addition the mobile application was developed as a proof of concept using the results of the algorithm.

This chapter explains each factors in the RFR-U model and its importance. It also discuss on how the model was built using the RFR-U factors.

## 4.2 RFR-U Model

The model uses passed data of three months of the intended user from the purchasing date. Then the data is further filtered using the RFR variables.

A rating for each item is calculated using RFR variables. Each variable is given a weight according to its significance. After the rating the probability of purchase for each item is calculated. Thereafter the result is further filtered by eliminating items which has the probability percentage less than 0.5.

The algorithm will output the predicted item list along with its predicted quantities and a similar item/s of interest.

### 4.2.1 Relevance

This parameter checks how relevant is an item to the purchasing date and time.

It was identified during the research that the users purchase different products during different times of the day and day of the week. As an example some people tend to buy rice during the lunch time every Friday as some do not bring lunch on Fridays.

This checks the relevance in terms of time and the day

a. Time

   This parameters checks how relevant is an item in the list for this time of the day. The weight table for the difference is given below. As this parameter could reduce the item list approximately by 50%, it is used as the first filtering criteria

| Time difference | weight |
|---|---|
| Purchasing time is between $\pm 30\ mins$ of item purchased time | 7 |
| Purchasing time is between $\pm 1.25\ hr$ of item purchased time | 5 |
| Purchasing time is between $\pm 2\ hr$ of item purchased time | 3 |

*Table 1: Weight table for time factor*

b. Day

   This checks how relevant is an item in the list for the day of the week. If the item matches with the day of the week the rating of the item will be increased.

Even though the result set can be drastically reduce, these parameter is not as significant as the recency parameter. Therefore the algorithm uses addition to increase the rating of the item.

### 4.2.2 Recency

This refers to recently purchased items.

It is likely that the user purchase the same item he/she bought recently. The reasons for this behavior could be an ongoing promotion, weather condition, current health situation etc.

This parameter compares the items purchase dates in-terms of the week of purchase. Since this is parameter is the most important parameter the algorithm uses multiplication to increase the rating. The weight table for the recency variable is given below.

| Date difference | weight |
|---|---|
| Item purchased date is within 1 week of the purchasing date | 10 |
| Item purchased date is within 2 week of the purchasing date | 8 |
| Item purchased date is within 3 week of the purchasing date | 5 |

*Table 2: Weight table of the recency parameter*

### 4.2.3 Frequency

This refers to number of times the user has purchased the item within the given time frame.

This is used as initial rating and the base for weight calculations. As an example if the certain item has been purchase 15 times during the past three months, then the frequency of the item is considered as 15. Likewise each item will have a frequency for each factor as well.

## 4.3 Variation of RFR-U model

Different variations of the model can be obtained by removing one or more variables from the model.

1. RFR model

   This model removes the user variable and the output is a list of items that all users would buy on the time of purchase. This predicts the types of items that customers would buy at a given time and date. It is useful for the store to plan their manufacturing.

2. RF model

   This model only considers the variable recency and frequency. This is outputs the items that users have recently purchased along with its probabilities. This is useful for the store to determine the fast moving items there by plan their menus and its quantities. Also this is useful to check if any on-going promotions have any effect.

3. RLF model.

   This model considers the relevance and frequency. The output of the model is a set of items that the user would buy at on a given time of the day. This is useful for the store to alter their menus.

## 4.4 Dashboard

The back-end dashboard illustrates the results of the RFR-U model and its variances. Below is a screenshot of the dashboard.

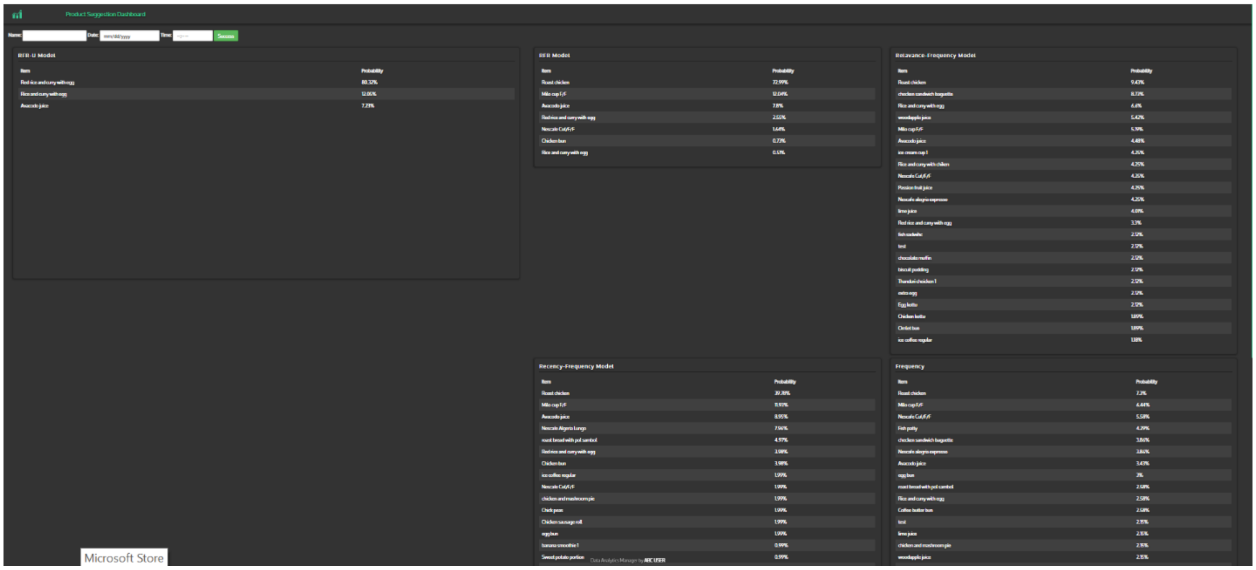



*Figure 17: Screenshot of the dashboard*

## 4.5 Mobile Application

The mobile application is developed as a proof of concept. It was developed as an Android application which supports 7 inch tabs. The application is named as “I-Checkout” to resemble its self-checkout functionality.

The user could enter his/her loyalty id and login to the system. Upon login, the user is directed to the home screen where a list of predicted items are displayed. The application uses the RFR-U algorithm to generate the said list. The result set of the algorithm is obtained using the services hosted in the back-end.

The user could tap on the icons to add preferred items to the food basket and using the checkout button a unique code can be generated. This code can be used for further processing such as payments.

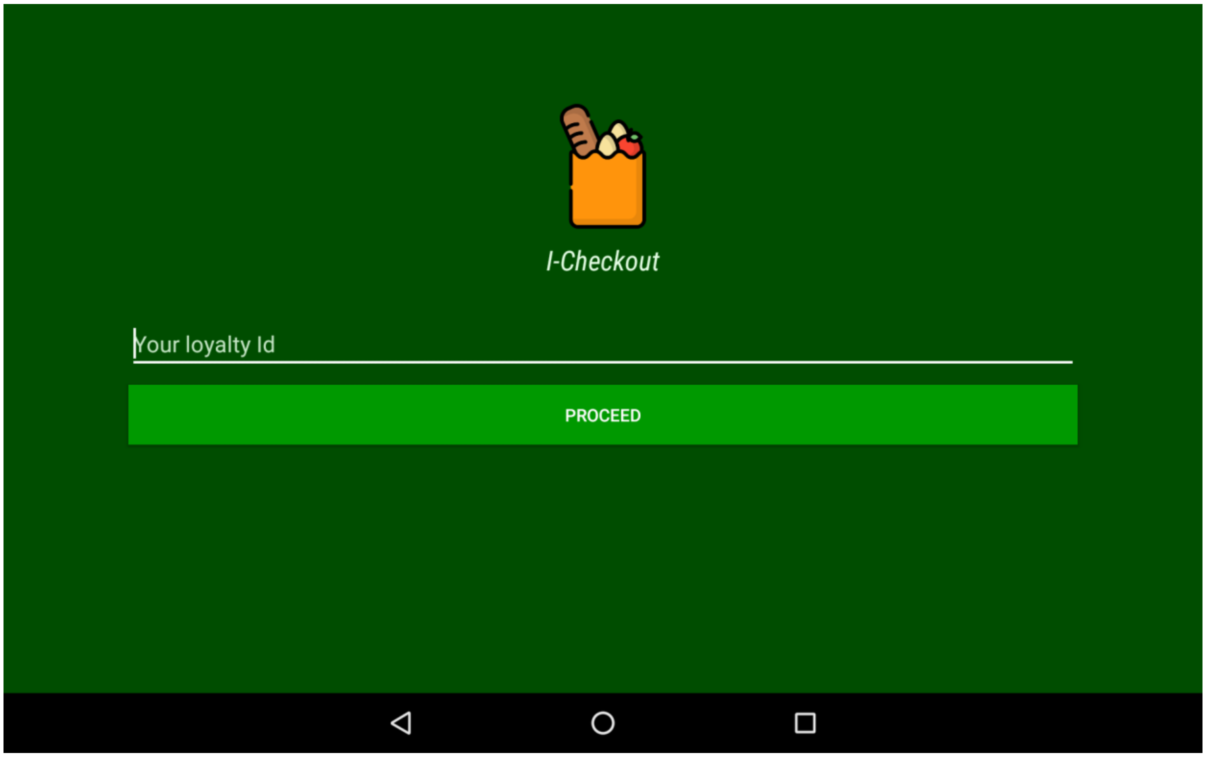


*Figure 18: Login screen of the POC app*

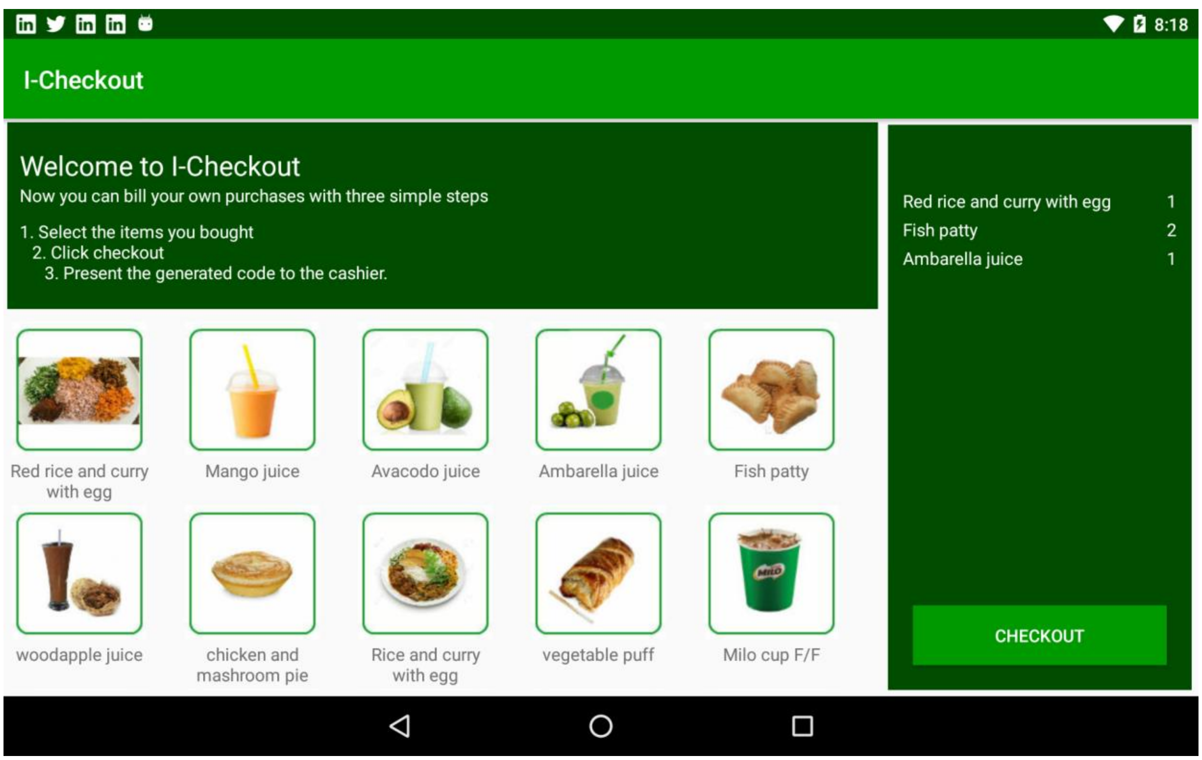


*Figure 19: Home screen of the POC app*

***

# Chapter 05

# Evaluation and Results

## 5.1 Introduction

The evaluation of the algorithm and the mobile application was done in both quantitative and qualitative approaches.

For quantitative approach different types of validation equations have been used to evaluate the algorithm and this chapter includes the test results of the evaluation. For the quantitate approach, feedback obtained from different users were used.

Fresh item purchasing bills of a selected set of users has been collected daily over a period of six months. The information in those bills have been inserted into a database and those data have been used as the data set for the research and evaluation.

The overall accuracy of the algorithm is around 83%

## 5.2 Quantitative Evaluation

Fresh bakery item purchasing bills were collected over a period of six months. As the algorithm uses historical data of three months to predict the next purchasing item list, the bills collected during the months of September, October, November and December were not used for the evaluation.

Gathering feedback from the users regarding the accuracy of the algorithm is an acceptable method. However depending completely upon the users' feedback on its correctness may not give an accurate picture about the accuracy of the algorithm. Further, to make a fair judgement it will be required to test many scenarios, such as, different days, different times and different users. This is a difficult and time consuming task. Therefore, a different approach had to be used to determine the accuracy of the algorithm. As a better approach, it was decided to use the gathered bills to evaluate the results.

The bills collected during the last three months (January, February, and March) were used as the testing data set. Since more than 100 bills are available it was able to evaluate the algorithm for different users and different time and date combinations.

A total of 25 bills were selected from the available bill collection which covers different user and date combinations. The user id, the purchased date and time in the bill was entered using the dashboard created and then the predicted item set was obtained using the algorithm.

The steps for the test is as follows;

1. Pick a bill from the bill set
2. Open the dashboard
3. Enter the user id, date and time in the dashboard
4. Click search button to generate results
5. Compare the results against the list in the bill
6. Enter the findings in the evaluation excel

The generated (Predicted) item set was compared against the actual item set in the bill. If the items predicted by the algorithm matches the items in the bill, it can be concluded that the algorithm has 100% accuracy. If few items are present it can be concluded that the algorithm is partially accurate. To calculate the accuracy percentage of the individual bills and the total accuracy, the following set of equations were used.

Let,

$P(n) = \ Correctly\ predicted\ number\ of\ items\ (within\ first\ n\ records)$

$N = Number\ of\ records\ in\ the\ bill$

1. **Rate of accuracy**

   The accuracy of the predicted items of a single bill

$$\boxed{Rate\ of\ Accuracy(n) = RA(n) = \frac{P(n)}{N} \times 100\%}$$

2. **Strict rate of accuracy**

   The algorithm may sometimes predict a list with many items. Therefore, it should be evaluated if the algorithm could correctly define the probability of purchase. This equation compares the items in the bill with the first five items which was predicted by the algorithm. Therefore this equation is used to check if the algorithm could correctly order the items predicted.

$$\boxed{Strict\ Rate\ of\ Accuracy\ (SRA) = \frac{P(5)}{N} \times 100\%}$$

3. **Weighted rate of accuracy**

   This equation is used to calculate the general accuracy of a single bill prediction

$$\boxed{Weighted\ Rate\ of\ Accuracy\ (WRA) = \frac{(\ 3\ \times RA(5)) + (\ 2\ \times\ RA(10)) + \ \big(RA(15)\big)}{6}}$$

4. **Average rate of accuracy:**

This equation is used to calculate the average accuracy of the complete set of bills. The accuracy can be calculate separately for strict accuracy and weighted accuracy.

$$Average\ rate\ of\ Accuracy = \sum_{i=1}^{m} \frac{SRA_i}{m} = \frac{1}{m} \sum_{i=1}^{m} \frac{P_i(5)}{N_i} \ ; \ m = number\ of\ bills$$

$$Overall\ Rate\ of\ Accuracy\ (Weighted) = \sum_{i=1}^{m} \frac{WRA_i}{m} \ ; \ m = number\ of\ bills$$

5. **Accuracy of a user:**

It is important to check the algorithms behavior for different users. This equation is used to separately calculate the average accuracy of a single user.

$$Average\ rate\ of\ Accuracy\ (U) = \sum_{i=1}^{m} \frac{SRA(u)_i}{m(u)} \quad ; \ m \rightarrow number\ of\ bills,$$
$$u \rightarrow selected\ user$$

$$Weighted\ Average\ rate\ of\ Accuracy\ (U) = \sum_{i=1}^{m} \frac{WRA(u)_i}{m(u)} \ ; \ m \rightarrow number\ of\ bills,$$
$$u \rightarrow selected\ user$$

## 5.3 Results of the Quantitative Analysis

The following table illustrates the test results obtained by the quantitative evaluation.

*Table 3: Test results*

| Bill date | User | N | P (5) | P(10) | P(15) | SRA / RA(5) | RA(10) | RA(15) | WRA |
|---|---|---|---|---|---|---|---|---|---|
| 07-02-2019 13:57 | 3 | 2 | 1 | 2 | 2 | 50 | 100 | 100 | 75 |
| 13/02/2019 12:38 | 3 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| 13/02/2019 08:46 | 3 | 2 | 2 | 2 | 2 | 100 | 100 | 100 | 100 |
| 07-02-2019 13:56 | 15 | 2 | 1 | 1 | 1 | 50 | 50 | 50 | 50 |
| 01-02-2019 12:07 | 1 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| 22-02-2019 09:46 | 2 | 6 | 4 | 4 | 4 | 66.666667 | 66.666667 | 66.6666667 | 66.666667 |
| 13-02-2019 17:28 | 15 | 3 | 2 | 2 | 2 | 66.666667 | 66.666667 | 66.6666667 | 66.666667 |
| 07-01-2019 14:10 | 3 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| 10-01-2019 11:58 | 13 | 2 | 2 | 2 | 2 | 100 | 100 | 100 | 100 |
| 10-01-2019 09:35 | 3 | 2 | 2 | 2 | 2 | 100 | 100 | 100 | 100 |
| 03-01-2019 09:55 | 3 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| 17-01-2019 09:03 | 1 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| 28-01-2019 17:04 | 2 | 3 | 2 | 2 | 2 | 66.666667 | 66.666667 | 66.6666667 | 66.666667 |
| 21-01-2019 12:55 | 1 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| 16-01-2019 08:35 | 3 | 2 | 1 | 2 | 2 | 50 | 100 | 100 | 75 |
| 29-01-2019 09:41 | 3 | 2 | 2 | 2 | 2 | 100 | 100 | 100 | 100 |
| 21-01-2019 08:17 | 3 | 2 | 2 | 2 | 2 | 100 | 100 | 100 | 100 |
| 24-01-2019 9:59 | 13 | 3 | 2 | 2 | 2 | 66.666667 | 66.666667 | 66.6666667 | 66.666667 |
| 14-01-2019 08:33 | 1 | 3 | 2 | 2 | 2 | 66.666667 | 66.666667 | 66.6666667 | 66.666667 |
| 26-03-2019 16:21 | 3 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| 13-03-2019 13:16 | 3 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| 21-03-2019 10:37 | 3 | 2 | 2 | 2 | 2 | 100 | 100 | 100 | 100 |
| 13-03-2019 08:56 | 3 | 2 | 2 | 2 | 2 | 100 | 100 | 100 | 100 |
| 05-03-2019 12"40 | 3 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| 08-03-2019 09:23 | 3 | 6 | 4 | 5 | 5 | 66.666667 | 83.333333 | 83.3333333 | 75 |
| 03-01-2019 17:17 | 15 | 3 | 3 | 3 | 3 | 100 | 100 | 100 | 100 |
| 30-01-2019 18:08 | 1 | 2 | 1 | 1 | 1 | 50 | 50 | 50 | 50 |
| 24-01-2019 13:45 | 1 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| 10-01-2019 14:10 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 08-01-2019 | 1 | 1 | 1 | 1 | 1 | 100 | 100 | 100 | 100 |
| **Average** | | | | | | **83.333333** | **87.222222** | **87.2222222** | **85.277778** |

Using the individually calculated accuracy rates, the overall accuracy of the algorithm was calculated. Below are the calculated accuracy levels.

- Average rate of Accuracy of the bills = 83.34%
- Weighted average rate of accuracy of the bills = 85.28%

The accuracy of the individual users were also calculated.

- User 1
    - Average rate of accuracy of the user 1(Strict) = 78.57%
    - Average rate of accuracy of the user 1 (weighted) = 78.57%

  However without the anomaly test record the accuracy stands around 88%
- User 2
    - Average rate of accuracy of the user 2(Strict) = 77.78%
    - Average rate of accuracy of the user 2 (weighted) = 77.78%
- User 3
    - Average rate of accuracy of the user 2(Strict) = 91.11%
    - Average rate of accuracy of the user 2 (weighted) = 96.42%
- User 15
    - Average rate of accuracy of the user 2(Strict) = 72.22%
    - Average rate of accuracy of the user 2 (weighted) = 72.22%

It was observed during the evaluation that the accuracy of the algorithm increases as the number of records increases. The accuracy drops when purchasing a completely new item which the user has not purchased before. This can be considered as an anomaly. Also there are situations where one person has bought items for few other people under the same bill. In such cases the accuracy gets less than 80%.

It is important to note that out of the 30 test cases listed above, 19 test cases have an accuracy of 100%. Considering the weighted average 22 test cases have an accuracy greater than 75%. Therefore it is clear that in most of the scenarios the algorithm has correctly predicted the next purchasing items.

## 5.4 Qualitative Evaluation

In addition to the quantitative evaluation a qualitative evaluation was also done to check if the research has achieved its objectives. A qualitative evaluation is important to evaluate the non-quantifiable aspects of the research such as user friendliness. This is also important to understand the different opinions of users, about the concept and the solution.

As the quantitative evaluation, a feedback form was given to eight individuals who purchase fresh items from the selected outlet on a daily basis. The individuals were given a chance to use the developed prototype system and provide their feedback on the system and the concept through the feedback form. The feedback form contains questions related to the accuracy of the results, satisfaction and other non-functional objectives of the solution. 'Google Forms' were used to create and distribute the feedback forms. The feedback form is attached in the appendices section for reference.

The summary of the collected responses is mentioned in the next section of this chapter.

## 5.5 Results of the Qualitative Analysis

Below graphs summarize the answers provided by the users for the feedback form.


*Figure 20: Answers for question 1*


*Figure 21: Answers for question 2*

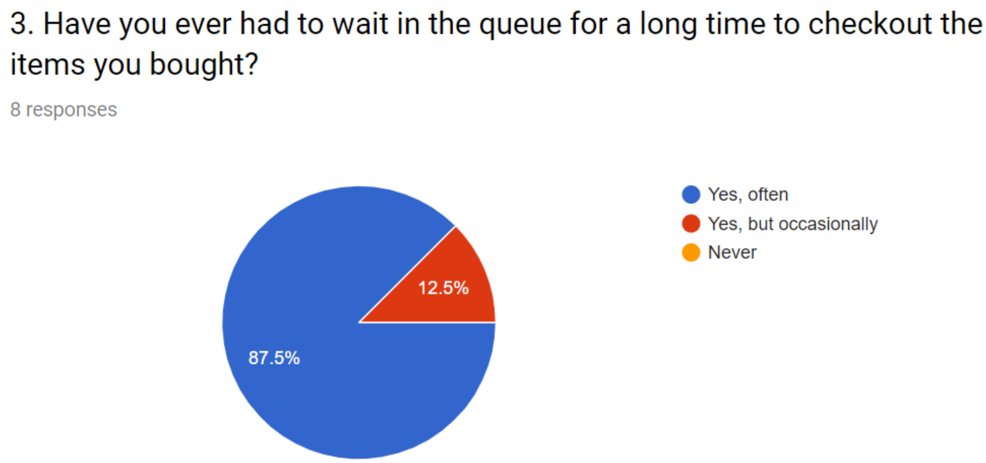



*Figure 22: Answers for question 3*

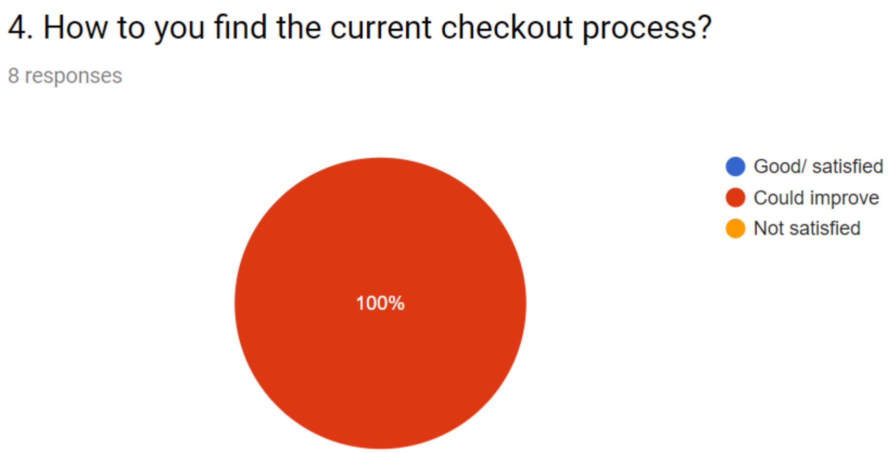



*Figure 23: Answers to question 4*

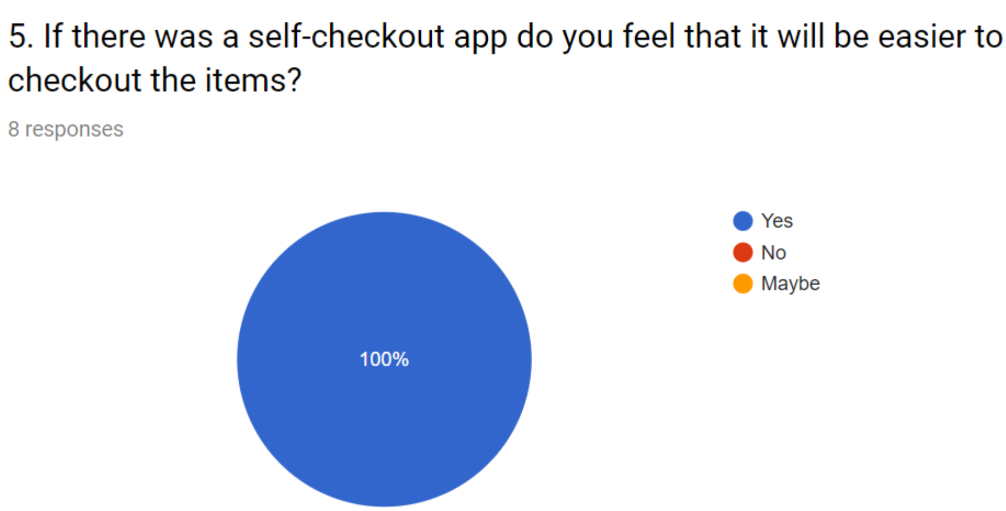



*Figure 24: Answers for question 5*

6. How do you find the proposed checkout application

8 responses

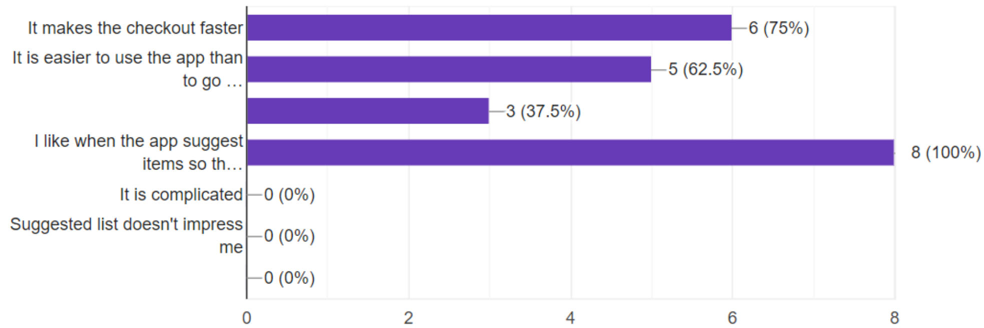



*Figure 25: Answers for question 6*

7. Does the proposed app make the checkout easier?

8 responses

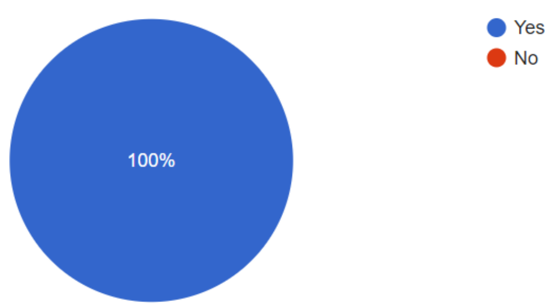



*Figure 26: Answers for question 4*

8. Does the app correctly show your preferred choices?

8 responses

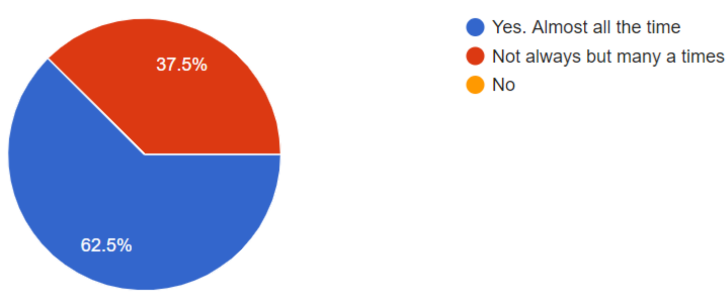



*Figure 27: Answers to question 8*

9.How satisfied are you with the results the application shows?

8 responses

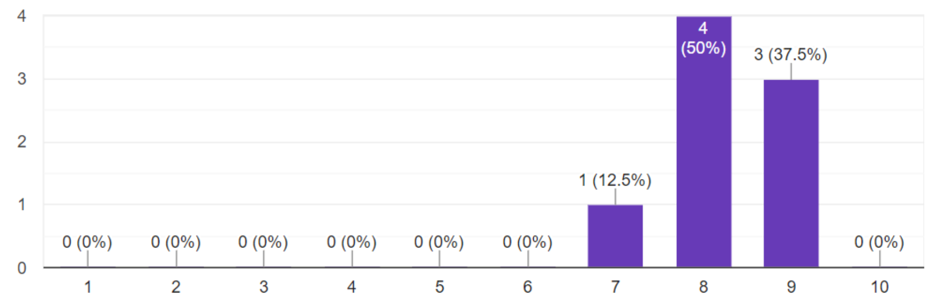



*Figure 28: Answers to question 9*

10. How satisfied are you with the overall application?

8 responses

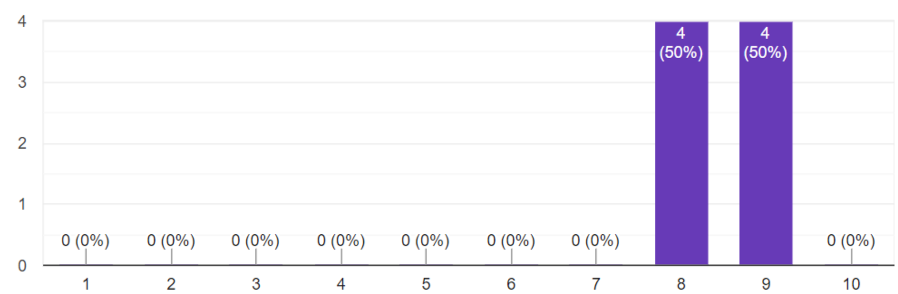



*Figure 29: Answers for question 10*

11. If the application was available in the store do you think this could reduce the queue size?

8 responses

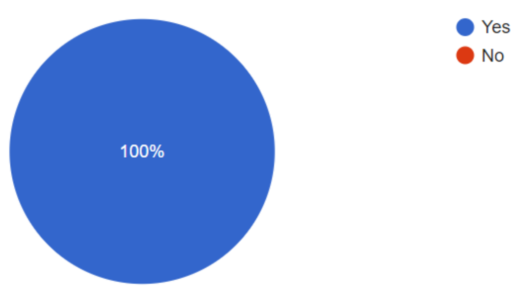



*Figure 30: Answers for question 11*

12. What are your suggestions for improvements?

8 responses

| |
|---|
| This Application can further be expanded to do all sorts of grocery shopping if the access was given to the DB of all the products and a payment gateway is integrated. |
| if the overall process of purchasing an item can be completed by the app it will be an added advantage. |
| If the payment facility is also incorporated into this app it is better. So we can do tge purchasing through the app itself rather than waiting in the queue to pay. |
| Its good if the app can show or filter according to the item I have already selected. |
| Sometimes App prompts the items I have not selected. |
| Introduce payment options with the app this will speed up the checkout process even faster. With this customers only need to collect the receipt from the check out counter. |
| Focus on a simplified design and towards usability |
| It would be better if you can integrate the payment process to the app through a payment gateway. As it will avoid the time spent at the cashier. |

*Figure 31: Answers for question 12*

As the graphs indicate a positive feedback has been received for the solution. All the users have agreed that the current checkout process can be improved and the proposed solution makes the checkout process easier. The satisfaction rate of the results produce by the algorithm stands above 80% and the satisfaction of the overall application and the concept also stands above 80%.

Most of the users liked the fact that the application filters and shows a predicted item list. Also users have mentioned that it is faster and easier to use the proposed approach than to use the existing billing mechanism.

The users were asked to give suggestion for improvements. Below are few points to note of the given answers.

1. Integrating a payment gateway, so that the customers will have the facility to pay through the application
2. Filter results as the customer add items to the food basket.

***

# Chapter 06

# Conclusion and Future work

## 6.1 Introduction

The industry trend is such that, the era of static UI is coming to an end. This is why the world today, talks about AI in UI. As the world becomes smarter, the user interfaces should also look smarter. The user interfaces will look smart, if they could display personalized content. Then the user will feel that the application knows who he/she is.

To make a personalized content, it is important to separately identify the user and his/her preferences. There are many data analysis techniques to classify and identify user behaviors. After identifying the behavior of the user, those algorithm can make predictions about the users future actions.

During the literature survey many techniques such as collaborative filtering methods, content based filtering methods, decision trees and matrix method were studied. But the studied scenario in the research requires a much simpler solution, so that the results could be obtained efficiently. Therefore a simple content, rule based and weighted rating method was used to implement the algorithm.

Through observation and data analysis different factors/variables which may directly define the patterns were identified. Time of purchase, date of purchase, day of purchase, frequency of the purchases were the important variables and according to the degree of significance of the variables, an algorithm was designed and developed to predict the next purchasing items set.

The algorithm uses the mentioned variables and heuristic weights to determine the result set. The algorithm outputs a set of items as, predicted next purchasing item set along with the probability of purchasing the item.

This algorithm was used in a self-checkout application which was designed to validate the concept. Using the results of the algorithm the application will display a personalized set of items for the user, making the application look and behave smarter.

## 6.2 Critical evaluation of the project

The main objective of the research was to design and implement an algorithm to predict the next purchasing item list of a given user, using his/her historical purchasing data. As mentioned in the literature review numerous methods and techniques are available to determine purchasing patterns of users. However, as the algorithm should generate results real time, a time efficient method had to be used to generate the results.

The RFR-U algorithm, which is elaborated in the thesis is able to generate a predicted next purchasing item set of a given user within seconds. During the research it was able to identify important variable that impact the purchases of the users. This is not only important for this research but will be important for future studies as well.

The accuracy rate of the algorithm is above 80%. The accuracy level is highly satisfactory as simple rule based methods have been used to design the algorithm. The feedback from the users have also been mostly positive. Hence it can be concluded that the solution is able to solve the problem mentioned in the research.

However, the algorithm is able to achieve the said accuracy level only if the customer is a regular customer. In addition there are scenarios where sometimes one customer would buy for several others as well. In such a scenario the accuracy level will be lesser than the mentioned accuracy level. As the algorithm uses simple techniques, the algorithm cannot predict anomalies.

## 6.3 Future work

The algorithm could be further improved by incorporating factors such as weather to generate results. As an example users would tend to buy hot beverages instead of cold ones on a rainy day. Therefore it is likely that the weather could influence the purchase. Therefore considering weather patterns when generating the result would give better accurate results.

The application can be made further personalized by filtering results according to the current selection. Many researchers suggest that customers are likely to purchase certain items together. Therefore using such mechanisms the application can be made more personalized and user friendly.

***

# References


[1] S. Tiwari, H.M. Wee, Y. Daryanto, "Big data analytics in supply chain management between 2010 and 2016: Insights to industries", [Online] Available: https://www.sciencedirect.com/science/article/pii/S0360835217305508 [Accessed July. 30, 2018].

[2] Y. Wang, L. Kung, T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations". [Online] Available: https://www.sciencedirect.com/science/article/pii/S0040162516000500 [Accessed July. 30, 2018].

[3] R. Iqbal, F. Doctor, B. More, S. Mahmu, U.Yousuf, "Big data analytics: Computational intelligence techniques and application areas". [Online] Available: https://www.sciencedirect.com/science/article/pii/S0040162517318498 [Accessed July. 30, 2018].

[4] G. Manogaran, D. Lopez , "Spatial cumulative sum algorithm with big data analytics for climate change detection ", Computers & Electrical Engineering, Volume 65, online] Available: https://www.sciencedirect.com/science/article/pii/S004579061730811X. [Accessed July. 30, 2018].

[5] D. V. den Poel, W. Buckinx, European Journal of Operational Research, Volume 166, Issue 2, [Online] Available: https://www.sciencedirect.com/science/article/abs/pii/S0377221704002875. [Accessed May. 24, 2018].

[6] H. Kim, G. Yang, H. Jung, S. Ho Lee, J. J. Ahn, "An Intelligent Product Recommendation Model to Reflect the Recent Purchasing Patterns of Customers", [Online] Available: https://link.springer.com/article/10.1007/s11036-017-0986-7. [Accessed May. 24, 2018]

[7] Predictive analytics. [online] Available: http://www.ijrbsm.org/pdf/v2-i11/5.pdf [Accessed September. 14, 2018]

[8] Hughes, A. M.. Strategic database marketing. Chicago: Probus Publishing Company. 1994.

[9] Z. Shi, Z. Wen, Jin Xia, " An Intelligent Recommendation System based on Customer Segmentation", International Journal of Research in Business Studies and Management, Volume 2, Issue 11

[10] Marcus C. A practical yet meaningful approach to customer segmentation[J]. Journal of consumer marketing, 1998, 15(5): 494-504.

[11] S. Lee, D. Lee, "A personalized channel recommendation and scheduling system considering both section video clips and full video clips", [Online] Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0199748 [Accessed September. 14, 2018]

[12] General Data Protection Regulation [online] Available: https://en.wikipedia.org/wiki/General_Data_Protection_Regulation [Accessed November. 14, 2018]

[13] GDPR Key Changes [online] Available: https://eugdpr.org/the-regulation/ [Accessed November. 16, 2018]

[14] M. Papagelis, D. Plexousakis, "Qualitative Analysis of User-based and Item-based Prediction Algorithms for Recommendation Systems", Engineering Applications of Artificial Intelligence 18 (2005)

[15] H. Lu, "Recommendations Based on Purchase Patterns", International Journal of Machine Learning and Computing, Vol. 4, No. 6

[16] M. Salehi, "An effective recommendation based on user behaviour: a hybrid of sequential pattern of user and attributes of product", Int. J. Business Information Systems, Vol. 14, No. 4

[17] Z. Zeng, "An Intelligent E-commerce Recommender System Based on Web Mining", International journal of business and management, Volume 4, Issue 7

# Appendices

1. The feedback form